RESEARCH ARTICLE

Technikvertrauen

Beiträge zur Technikfolgenabschätzung jenseits von Akzeptanz und Akzeptabilität?

Sebastian Weydner-Volkmann, *Institut für Philosophie I, Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, DE*
(sebastian.weydner-volkmann@ruhr-uni-bochum.de) *0000-0003-3948-4770*

Zusammenfassung • Der Beitrag lotet aus, inwiefern über den Begriff „Technikvertrauen" komplementär zu den in der Technikfolgenabschätzung bereits etablierten Begriffen „Akzeptanz" und „Akzeptabilität" ein konzeptueller Beitrag für eine ethische Technikbewertung geleistet werden kann. Es wird gezeigt, dass gerade für Digitaltechniken Aspekte der Angriffssicherheit besser adressiert werden können, weil hier an die Begrifflichkeiten der IT-Sicherheitsforschung angeschlossen werden kann. Zudem erlaubt „vertrauenswürdige Technik" eine bessere Einbeziehung von Laienperspektiven, da ein rational begründetes Vertrauen im Sinne von Risikoerwartungen interpersonell durch Expertisen vermittelt werden kann. Insbesondere für die Bewertung von Digitaltechniken kann „Technikvertrauen" somit eine Lücke zwischen Akzeptanz und Akzeptabilität schließen.

Trust in technology. Ethical contributions to technology assessment beyond acceptance and acceptability?

Abstract • *This article explores the potential for "trust in technology" to make a productive conceptual contribution to the ethical evaluation of technology, complementing the concepts of "acceptance" and "acceptability" already established in technology assessment. It shows that for digital technologies in particular, "trust" can better address aspects of security against attacks as it allows to integrate concepts of IT security. Furthermore, "trustworthy technology" allows for a better inclusion of lay perspectives, since rationally justified trust in the sense of risk expectations can be mediated interpersonally by experts. Especially for the evaluation of digital technologies, "trust in technology" can thus bridge a conceptual gap between acceptance and acceptability.*

Keywords • *trust, technology ethics, digital technology, IT security, echnology assessment*

https://doi.org/10.14512/tatup.30.2.53
Received: Feb. 22, 2021; revised version accepted: May 19, 2021;
published online: Jul. 26, 2021 (non-blind peer review)

Vertrauen in Technik?

Zum Begriff ‚Technikvertrauen' einen konzeptuellen Beitrag für die Technikfolgenabschätzung (TA) leisten zu wollen, ist erklärungsbedürftig. Zwar ist der Begriff durch politische Initiativen rund um Digitalisierung, *Machine Learning* und Künstliche Intelligenz (KI) zuletzt stärker in den Fokus gerückt (AI HLEG 2019) und kann zunehmend auch in der europäischen und nationalen Forschungsförderung angetroffen werden. Allerdings spielt der Begriff für die Technikbewertung bislang keine nennenswerte Rolle. Auch disziplinär ist ‚Technikvertrauen' in Philosophie und Ethik einigermaßen umstritten: Als „eigentliche", prototypische Domäne des Vertrauens gelten zwischenmenschliche Beziehungen (Budnik 2016, S. 68–70; Köhl 2001, S. 131). Es sei auch noch kein plausibler Begriff des Vertrauens formuliert worden, der nicht interpersonell geprägt ist; beim ‚Technikvertrauen' handle es sich vielmehr entweder um eine alltagssprachliche Verwechslung mit dem Phänomen des ‚Sich-verlassens-auf' (Verlässlichkeit, *reliability*) oder um eine verkürzte Ausdrucksweise für Vertrauen in die jeweiligen Konstrukteure der Technik (Hartmann 2010, S. 20). Bisweilen wird beim Technikvertrauen gar „begrifflicher Unsinn" aus dem ‚Werbesprech' der Technikindustrie ausgemacht (Metzinger 2019). So gesehen scheint es wenig Grund für das Anliegen dieses Artikels zu geben, für ethische Beiträge zur TA auszuloten, inwiefern über das Konzept ‚Technikvertrauen' ein konzeptueller Beitrag geleistet werden kann.

Allerdings gibt es in den letzten Jahren Entwicklungen, die neue Perspektiven auf Technikvertrauen erschließen lassen und über die Anschlussmöglichkeiten an die TA-Debatten plausibel werden: Angesichts der Vielzahl sich oftmals widersprechender philosophischer Vertrauenskonzeptionen wird vermehrt gefordert, den Zugriff differenziert über bereichsspezifische Kriterien der Vertrauens*würdigkeit* oder auch vom *Wert* des Vertrauens her zu führen (Budnik 2016, S. 70 f.; Jones 2012). Noch weiter gedacht soll die Vertrauenskonzeption gar insgesamt an den konkreten Klärungsabsichten ausgerichtet werden: "An alternative theoretical approach is […] to ask not what our concept is, but what it ought to be, if we wish the notion to do useful con-

ceptual work. Proposed accounts are to be evaluated against our legitimate purposes, which opens up the possibility that different accounts might suit different purposes." (Jones 2020, S. 6 f.)

Um zu zeigen, dass ‚Technikvertrauen' in diesem letzten Sinne *produktiv* als intellektuelles Werkzeug für die Technikbewertung konzipiert werden kann, darf das alltagssprachlich akzeptierte semantische Feld des Vertrauens natürlich nicht gänzlich verlassen werden. Auch wenn in einer solchen Konzeption keine „prototypischen Vertrauensverhältnisse" zum Ausdruck kommen sollten, muss ein produktives Konzept von Technikvertrauen intuitiv noch *als Vertrauensverhältnis verständlich bleiben*. Andernfalls würde die Begriffsbildung letztlich der Beliebigkeit preisgegeben.

Auch dürfen bereits etablierte TA-Konzepte nicht ohne weiteres zugunsten einer bloßen Reformulierung unter dem Vorzeichen des Vertrauens verworfen werden. Um *produktiv* zu sein, muss ein solcher Vertrauensbegriff es daher erlauben, im Rahmen einer ethischen TA *neue Aspekte oder ein vertieftes Verständnis beim bewertenden Zugriff auf Techniken zu erschließen.* Entsprechend müssen sich die Stärken eines in diesem Sinne konzipierten ‚Technikvertrauens' gegenüber etablierten Konzepten argumentativ ausweisen lassen. Für eine Technikbewertung können hier insbesondere die Begriffe ‚Akzeptanz' und ‚Akzeptabilität' genannt werden (Gethmann und Sander 1999, S. 146).

Im Folgenden untersuche ich unter Einbeziehung der Forschungsliteratur, inwiefern ‚Technikvertrauen' relevante Aspekte für eine ethische Technikbewertung erschließen lässt. Dabei prüfe ich auch, ob diese evaluativen Konzeptionen über die bewährten Begriffe ‚Akzeptanz' und ‚Akzeptabilität' hinausgehen. Für legitimationsstiftende Aspekte wie auch für Verlässlichkeitsfragen lässt sich dies verneinen. Gerade für Digitaltechniken zeige ich in Abschnitt 4 jedoch, dass Erwartungshaltungen zur Angriffssicherheit besser berücksichtigt werden können, weil hier an die Begrifflichkeiten der IT-Sicherheitsforschung angeschlossen werden kann. Zudem zeige ich, dass ‚vertrauenswürdige Technik' es erlaubt, eine auch für Laien rational begründete Erwartungshaltung zu Technikrisiken in den Blick zu nehmen.

Legitimation durch Vertrauen

Als evaluativer Begriff einer Technikethik kann die Frage nach dem Vertrauen zunächst einmal als Frage danach angegangen werden, ob einer eingesetzten Technik *de facto* von Anwendern oder Betroffenen subjektiv vertraut wird. Dies kann insbesondere im Sinne einer demokratischen Legitimation oder einer individuellen Einwilligung relevant werden, speziell mit Blick auf diejenigen, die vom Technikeinsatz dann auch betroffen sind. Der Begriff ‚Technikvertrauen' übernimmt hierbei eine sehr ähnliche argumentative Funktion wie der in der TA schon länger etablierte Akzeptanz-Begriff (Jaufmann 1999). Darüber hinaus wurde Technik-Akzeptanz besonders in der Technikgestaltung fruchtbar gemacht:

> „Die Idee akzeptanzorientierter Technikgestaltung bestand darin, die angenommene Technikakzeptanz bereits in der Technikentwicklung zu berücksichtigen. Durch prospektive Untersuchungen sei herauszufinden, welche Technik (einschließlich ihrer Risiken und sonstiger Nachteile) faktisch wohl akzeptiert würde." (Grunwald 2005, S. 55)

Ganz offenbar können diese über ‚Technikvertrauen' angesprochenen Aspekte sehr gut und differenziert über den Akzeptanzbegriff adressiert werden. Im Bereich der Medizintechnik mag faktisch-empirisches Vertrauen z. B. die Frage adressieren, unter welchen Bedingungen eine Ärztin oder ein Patient sich für den Einsatz einer noch nicht etablierten Behandlungsmethode entscheiden. ‚Vertrauen' wie auch ‚Akzeptanz' verweisen hier konzeptuell auf ein allgemeineres ‚Sich-verlassen-auf' und somit auf subjektive Einschätzungen zu Betriebssicherheit, Ausfallwahrscheinlichkeit, Präzision etc. Unser faktisches Technikvertrauen wird dabei darin manifest, dass wir uns für die Verwendung einer Technik entscheiden, uns dabei ein Stück weit auf sie verlassen und ihr ggf. den Vorzug gegenüber alternativen Techniken geben.

Neben legitimatorischen Fragen wird dieses Vertrauen normativ insbesondere in Vergleichen alternativer Technik-Optionen relevant: Ist das Vertrauen in die Leistungsfähigkeit einer innovativen Behandlungs-Technik angemessen oder aber im Vergleich zu anderen Techniken zu stark bzw. zu schwach ausgeprägt? Hierüber lassen sich dann etwa Aufklärungs- und Informationspflichten der behandelnden Ärztin bzw. des jeweiligen Herstellers ethisch begründen (Nickel 2011, S. 385 f.). Um einen solchen Vergleich normativ zu begründen, bedarf es jedoch gemeinsamer Bezugspunkte, die das Vertrauen mit Blick auf die Leistungsfähigkeit von Techniken in ein Verhältnis setzten können. Hier wird die Frage aufgeworfen, inwiefern sich dieses ‚Sich-verlassen-auf' rational begründen lässt – womit sich der Blick weg vom *de facto* Vertrauen und hin zur (rational begründeten) Vertrauens*würdigkeit* wendet.

Vertrauenswürdigkeit als Verlässlichkeit

Im Kontext von Mensch-Maschine-Interaktionen fällt der Begriff ‚Verlässlichkeit' bzw. *'reliability'* häufig, um den Begriff ‚Technikvertrauen' zu explizieren (Lampe und Kaminski 2019; Nickel 2011, S. 359). Als Bezugspunkt für die dann angesprochene Vertrauens*würdigkeit* spielen Verlässlichkeitsmaße zu Betriebssicherheit etc. eine zentrale Rolle. Dabei erlaubt die rationale Begründung von Vertrauensverhältnissen gerade auch eine *Kritik des faktischen Vertrauens*. Der Bezug auf Verlässlichkeitsmaße erschließt deshalb eine reflexive Perspektive auf emotional oder habituell motivierte Vertrauensverhältnisse zwischen Mensch und Technik (Ropohl 2010, S. 118 f.). Eine für die TA produktive Konzeption von Technikvertrauen muss diese Verlässlichkeitsmaße deshalb als Teil der rationalen Begründung in zentraler Weise berücksichtigen.

‚Vertrauenswürdigkeit' schließt somit überaus unzuverlässige Technik ebenso aus, wie im zwischenmenschlichen Bereich eine notorische Unzuverlässigkeit oder Inkompetenz (O'Neill 2018, S. 294). Man mag mangels Alternativen zwar auch unzuverlässige Technik nutzen, eine vertrauens*würdige* Technik muss jedoch einerseits die anvertrauten Aufgaben einigermaßen verlässlich erfüllen können und sie darf andererseits beim Einsatz keine übermäßigen Störungen im Sinne negativer Folgewirkungen erzeugen. Solche negativen Folgewirkungen, mit denen sich die TA ja seit Beginn intensiv beschäftigt, lassen sich als Risiken im weiteren Sinne fassen. Verlässlichkeitsmaße können so als Ausfall-, Unfall- oder Nutzungs-Risiken von Techniken mitberücksichtigt werden und korrespondieren insbesondere mit Anforderungen zur Betriebssicherheit (*safety*).

Im Technikvertrauen wird ein positiver Ausgang der Techniknutzung erwartet – trotz des Wissens um Risiken.

Normativ ist aber noch ungeklärt, was in einer konkreten Anwendung dann als „verlässlich genug" gelten darf, um Vertrauen rational zu begründen. Wie Kaminski (2010, S. 219–226) herausarbeitet, besteht hier auch kein direkter Zusammenhang zwischen faktischem Vertrauen und quantitativ bestimmbaren Risiken: Vertrauen ist keine Reaktion auf konkrete Risikowerte, sondern kann als *Erwartungshaltung* begriffen werden, in der man davon ausgeht, dass die Technik wie gewünscht funktionieren wird, sich das vorhandene Schadenspotential also *schlicht nicht manifestieren wird*:

> „Technologien bringen Risiken mit sich. Wird diesen vertraut, dann werden sie als eigentlich risikolos betrachtet. Es handelt sich dann mehr um prinzipielle, abstrakte Bedenken, die aber auf die jeweilige Technologie selbst nicht zutreffen. Vertrauen in Technik weist stets diese Form risikolosen Risikos auf […] Wer vertrauensvoll mit dem Auto oder im Internet unterwegs ist, negiert damit nicht, dass prinzipiell Risiken bestehen, nur erscheinen sie situativ nicht gegeben. Stets wird das mit Technik einhergehende Risiko als prinzipielles, situativ aber nicht relevantes betrachtet – sofern vertraut wird." (Kaminski 2010, S. 225 f.)

Im Technikvertrauen, so Kaminskis Pointe, schlägt die Erwartungshaltung gegenüber der Technik um: Wir erwarten einen positiven Ausgang in der Techniknutzung, obwohl wir durchaus wissen, dass Risiken bestehen, es also auch anders ausgehen könnte. Nur deshalb kann unser Vertrauen überhaupt *enttäuscht* werden – eben weil unsere Erwartung im Handeln eine andere war. Würde der ebenfalls mögliche negative Ausgang im Vertrauen nicht ‚ausgeklammert', so würde im positiven wie im negativen Ausgang etwas eintreten, was Teil unserer Erwartungshaltung war – wir hätten also gar keinen Grund zur Enttäuschung. Ganz ähnlich schreibt auch Ropohl (2010, S. 121), dass „die Gewissheit des Vertrauens als uneingeschränkt positive Erwartung erlebt wird".

Unter Verweis auf Luhmann legt Kaminski (2010, S. 255) dar, dass hierin auch ein funktionaler Wert des Vertrauens liegt, nämlich in der Reduktion von erlebter Komplexität: Im Vertrauen reagieren wir auf Risiken im Sinne einer ‚Ausklammerung' möglicher (negativer) Zukünfte. Eine lähmende Unsicherheit angesichts des ungewissen Ausgangs einer Techniknutzung kann so überwunden und es können neue Handlungsoptionen eröffnet werden. Kaminski (2010, S. 256) betont hierbei eine Kontinuität von Vertrauen in Personen zu Vertrauen in Technik: Ein umfassender Entzug des Vertrauens in die uns umgebende Technik sei zwar prinzipiell ebenso denkbar wie ein Entzug des Vertrauens in alle Mitmenschen, aber mit Blick auf die für unseren Alltag „nötige" Reduktion von Komplexität praktisch nachrangig.

Nach Kaminski (ibid., S. 195 f., 263 f.) ‚versinken' Alltagstechniken in eine *Vertrautheit*, durch die sie sich einer hinterfragenden Thematisierung ihrer Verlässlichkeit entziehen. Die Vertrautheit ist somit vom *Technikvertrauen* abzugrenzen, bei dem es um einen spezifischen Umgang mit Nichtwissen und Risiken geht. Entsprechend stellt sich vor allem für *neue* Techniken die Frage, wann es rational angemessen ist, den Sprung zu wagen und einer Technik zu vertrauen, sie also für vertrauens*würdig* zu befinden. Gerade weil es keinen direkten Zusammenhang zwischen Verlässlichkeitsmaßen bzw. Risiken einerseits und dem Technikvertrauen andererseits gibt, kann auch nicht technikinhärent geklärt werden, was es im Sinne einer normativ argumentierenden TA bedeutet, eine Technik als *verlässlich genug* gegen Störungen zu bezeichnen. Über die Hinterfragung der Vertrauenswürdigkeit im Rahmen einer TA wird somit eine Entscheidungsdimension angesprochen, die sich neben den Verlässlichkeits- bzw. Risikomaßen immer auch auf Wertvorstellungen beziehen muss, z. B. mit Blick auf die Zumutbarkeit von Risiken.

Auf diese notwendig wertbezogene Dimension risikobehafteter Entscheidungen ist in der TA-Debatte häufig genug hingewiesen worden. Konkrete ethische Begründungen für die Zumutbarkeit von Nebenfolgen werden innerhalb der TA dabei oft über den Begriff der *Akzeptabilität* eingebracht – gerade auch in Abgrenzung zu ihrer faktischen Akzeptanz (Grunwald 2005, S. 55; Weydner-Volkmann 2018, S. 33). Anders gesagt: Technikvertrauen klammert mögliche negative Zukünfte aus und deutet risikobehaftete Technik in unserer Erwartungshaltung in risikolos einsetzbare Technik um. Deshalb wirft die explizite Frage nach der Vertrauens*würdigkeit* von Technik auch ein breites Spektrum an normativen Fragen nach der *Akzepta-*

bilität auf. In diesem Sinne verweist ‚vertrauenswürdige Technik' in einer ethisch argumentierenden Technikbewertung darauf, eine Unterscheidung zwischen naivem (emotional oder habituell geprägtem) und rational begründetem Technikvertrauen zu machen.

Es scheint dabei, so ein erstes Zwischenfazit, als ließe sich diese Unterscheidung recht gut über die Doppelung der in der TA bereits etablierten Begriffe Akzeptanz und Akzeptabilität fassen, nämlich als eine Akzeptanz bei normativ unzureichend begründeter Akzeptabilität (naives Technikvertrauen) und bei gut begründeter Akzeptabilität (begründetes Technikvertrauen). Ich werde hierauf zurückkommen.

Täuschung und Vertrauenswürdigkeit

In den Debatten zur Vertrauenswürdigkeit von Technik wird ein Aspekt nur selten thematisiert, der im zwischenmenschlichen Bereich sogar als „Lackmustest für Vertrauen" in Abgrenzung zur bloßen Verlässlichkeit gilt (Budnik 2016, 107): So fühle man sich *ent*täuscht, wenn eine Technik nicht gut funktioniere oder sich eine Person am Ende doch als unzuverlässig herausstelle, aber man fühle sich eben nicht *ge*täuscht. Dabei wird der mögliche *bewusste Missbrauch* als Merkmal für Vertrauensverhältnisse hervorgehoben, was einen kompetenten Akteur voraussetzt. Entsprechend erklärt sich, warum dieser Aspekt des Vertrauensbegriffs (jenseits spekulativer Überlegungen zu ‚starker KI') in der Technikdebatte bislang kaum in den Blick gerückt ist.

Als Ausnahme darf Clemens Cap gelten, der versucht, vertrauenswürdige Technik als ein informationstechnisches System zu deuten, das sich unter anderem „auch den weitergehenden, abstrakteren Zielen und Interessen seines Besitzers entsprechend [verhält]" (Cap 2015, S. 115). Dabei beschreibt er das „Handeln" eines informationstechnischen Systems gegen die Interessen der Nutzer: „Emotional vermitteln sie dem Endanwender […], dass ein System, das eigentlich zu ihrer Unterstützung gedacht war, sich gegen sie wendet. Der Nutzer kann das als ‚Verrat' werten und wird es nicht mit Vertrauen assoziieren" (Cap 2015, S. 116–118). Cap hat an dieser Stelle insbesondere Funktionen im Blick, die Hersteller, Verkäufer oder Administratoren von informationstechnischen Systemen bewusst für Ihre eigene Zwecke nutzen (etwa um Nutzerprofile aufzubauen). Solche technischen Funktionen können dann zwar technisch sehr verlässlich umgesetzt sein, aber sie funktionieren eben nicht im Sinne der Ziele und Interessen des Endanwenders, sondern dienen dem Hersteller, Diensteanbieter, etc.

Cap (2015, S. 109) erwähnt zudem einen weiteren Typus des „Verrats" an Nutzerinteressen, nämlich bewusste Angriffe auf informationstechnische Systeme, deren stärkere Berücksichtigung bei Weber et al. (2020 a, S. 32) explizit als Desiderat für die TA herausgestellt werden. Neben dem bisher diskutierten Aspekt der Verlässlichkeit rückt hierbei die *informationstechnische Sicherheit* in den Vordergrund, die in der Informatik als eigenständiger Anforderungsbereich aufgefasst wird (Weber et al. 2020 a, S. 30). Verlässlichkeit (*reliability*) gilt dabei in der Informatik als Fähigkeit eines Systems, bis zu einem gewissen Zeitpunkt störungsfrei zu arbeiten (Eusgeld et al. 2008, S. 59) – also die Robustheit eines Systems gegen scheinbar „zufällig" auftretende Fehler, deren Eintrittswahrscheinlichkeit sich im Idealfall statistisch berechnen lässt und über die man unterschiedliche Lösungen hinsichtlich der Fehlerwahrscheinlichkeit vergleichen kann. Im Gegensatz zu solchen auf Betriebssicherheit (*safety*) bezogenen Fragen bezieht sich ‚Vertrauen' in der Informatik insbesondere auf Aspekte der Angriffssicherheit (*security*) – gängige Konzepte in der IT-Security sind z. B. *root of trust*, *chain of trust*, oder auch *trusted fab* (GlobalPlatform 2018; Weber et al. 2020 a, S. 33). Bewusste Angriffe lassen sich, im Gegensatz zu Fehlfunktionen, aber grundsätzlich nicht statistisch erfassen; so wird z. B. eine Angreiferin bewusst genau jene *höchst selten* auftretenden Fehlerbilder provozieren, die sie dann *verlässlich* für ihre eigenen Zwecke ausnutzen kann. Als Zielvorstellung der Methoden der IT-Security wird dabei im Allgemeinen auf drei Aspekte rekurriert, nämlich auf 1) die Integrität, 2) die Vertraulichkeit und 3) die Verfügbarkeit der Datenverarbeitung (Weber et al. 2020 a, S. 30).

Man mag hier einwenden, dass es ja nicht die Technik ist, die den Endanwender täuscht. Der eigentliche Akt der Täuschung wird hier doch eher vom Hersteller vollzogen, der eine geheime Zugriffsmöglichkeit (*backdoor*) einbaut, oder eben von der Angreiferin, die über eine Sicherheitslücke das System eines Endanwenders übernimmt. Diesem Einwand ist zunächst wenig entgegenzusetzen, doch tritt hier ein Aspekt des Technikvertrauens in den Blick, der für eine *normative Bewertung von Digitaltechniken* begrifflich ebenso sinnvoll erfasst werden muss, wie der Aspekt der Verlässlichkeit: Dient ein Gerät tatsächlich den Interessen des Endanwenders, oder wurde das Gerät kompromittiert? Dabei liegt die Täuschung ja gerade in der bewusst provozierten, falschen Erwartungshaltung: Das Gerät tut gerade nicht, was ich erwarte, aber dies ist kein Fall von *ent*täuschender technischer Unzuverlässigkeit, sondern, im Gegenteil, das Ergebnis einer absichtlichen Täuschung oder Verschleierung.

Für den Endanwender wäre *Misstrauen* hier offenbar das angemessenere Verhältnis – doch gegenüber wem bzw. gegenüber was? Geht man an dieser Stelle den naheliegenden Schritt, das Misstrauensverhältnis symmetrisch zum eigentlichen Akt der Täuschung aufzufassen (der Endanwender wird von der Angreiferin getäuscht, also sollte er der Angreiferin misstrauen, nicht seinem IT-System), so würde man letztlich an den Gegebenheiten der digitalen Praxis vorbei argumentieren. Denn es ist wahrscheinlich, dass *Angreiferin und Endanwender keinerlei persönliche Beziehung pflegen* – eventuell ist letzterer lediglich einer von vielen Nutzern, deren Systeme die Angreiferin übernimmt und deren persönliche Daten sie zum Zweck der Erpressung dann verschlüsselt. Das angemessene Vertrauensverhältnis des Endanwenders ist also nicht ein gezieltes Misstrauen gegenüber der Angreiferin, sondern ein *allgemeines Misstrauen* in der Nutzung des IT-Systems. Und in diesem Sinne kor-

respondiert die der Situation angemessene *Erwartungshaltung* auch nicht mit der Vertrauenswürdigkeit der Angreiferin, sondern mit der Robustheit der Technik gegen derartige Angriffe ganz allgemein.

Im Gegensatz zur „analogen" Welt, wo solche Angriffsszenarien für die allermeisten Menschen vernachlässigt werden können – wer überprüft schon sein Fahrrad vor der Fahrt auf Manipulation? – betreffen Cyberangriffe aufgrund der hochgradigen Vernetzung heute nahezu alle Menschen (Weber et al. 2020 b): Angriffe auf digitale Systeme sind global und mit geringem Aufwand *en masse* durchführbar, das Risiko entdeckt zu werden überschaubar; die Gesellschaft sieht sich aktuell mit Angriffen durch Hersteller, Dienstleister, Geheimdienste und kriminelle Hacker konfrontiert.

Technikmisstrauen bezieht sich nicht auf Personen, sondern auf die allgemeine Nutzung des IT-Systems.

Gerade für die immer zentraler werdenden Digitaltechniken eröffnet ein Rückgriff auf den Vertrauensbegriff über das Konzeptpaar ‚*security*' und ‚*trust*' für eine Technikbewertung somit einen Anknüpfungspunkt, um die höchst relevanten Forschungsdebatten in der IT-Sicherheit einzubinden. Methodisch lässt sich dies stimmig über die Erwartungshaltung der Endanwender, der Betroffenen, der interessierten Öffentlichkeit, etc. einlösen. Alternativ könnte man freilich auch versuchen, den Begriff der Akzeptabilität um den Aspekt der Angriffssicherheit zu erweitern; allerdings muss methodisch dann auch geklärt werden, inwiefern Angriffe z. B. im Sinne einer Zumutbarkeit von Risiken für eine ethische Technikbewertung sinnvoll erfasst werden.

Das Problem der Vermittlung von Vertrauen

Blickt man auf die vorherigen Abschnitte, so scheint im Rahmen einer ethischen TA neben den etablierteren Begriffen der Akzeptanz und der Akzeptabilität wenig Raum zu bleiben für eine produktive Rolle des Konzepts ‚Technikvertrauen' – zumindest dann, wenn es gelingt die relevanten Aspekte von ‚*security*' und ‚*trust*' über das Konzept ‚Akzeptabilität' einzubinden (was keineswegs ausgemacht ist). Immerhin scheinen die wesentlichen ethisch problematischen Aspekte überall dort gelöst, wo eine legitimierende Akzeptanz auf eine rational begründete Akzeptierbarkeit von Technik trifft, bzw. dort, wo eine Ablehnung auf das Urteil der Inakzeptabilität trifft.

Auf den zweiten Blick wird allerdings ein Vermittlungsproblem offenbar: Während Akzeptanz als politische Legitimation durch (Betroffenheits-)Öffentlichkeiten fungiert, wird Akzeptierbarkeit in der Regel als eine Eigenschaft von Technik innerhalb von bestimmten Einsatzszenarios gedacht. Das Vermittlungsproblem besteht dabei darin, dass sich eine Akzeptierbarkeit letztlich nicht nur im Expertenurteil (z. B. als Teil einer TA) rational ausweisen lassen muss, sondern gerade auch *im Urteil derjenigen, deren Akzeptanz eine legitimierende Funktion für Technikentwicklung und Technikeinsatz entfalten soll* – also in den allermeisten Fällen auch im Urteil von Laien. Wäre die faktische Akzeptanz für die Legitimation einer Technik, deren Risiken als zumutbar ausgewiesen wurden, für sich genommen hinreichend, so spräche nichts dagegen, diese auch über eine irreführende, emotionale Werbekampagne und Falschinformationen „herzustellen".

Grunwald (2005, S. 55, 58) beleuchtet zwar ein ähnliches Problem, nämlich die Nichtakzeptanz von sehr ungleich verteilten Risiken, und schlägt hierzu vor, statt auf akzeptierte *Ergebnisse*, auf akzeptierte *Verfahren* zur Ergebnisfindung zu setzten (Prozesslegitimität), etwa über Planfeststellungen mit Bürgerbeteiligung. Speziell bei Digitaltechniken scheint das Problem aber oft anders gelagert zu sein: Zum einen betreffen die Techniken hier häufig die Gesellschaft in ihrer Breite, teilweise gar als Infrastruktur, deren „akzeptierte" Nutzung zudem häufig für ein Gelingen der Technik Voraussetzung ist (z. B. die Corona-App, De-Mail, die elektronische Gesundheitsakte). Zum anderen kommt es, wie oben ausgeführt, aufgrund der Vernetzung viel stärker darauf an, eine je *individuelle Angriffssicherheit auch gegen staatliche Akteure und gegen die Betreiber und Entwickler der jeweiligen digitalen Dienste* zu gewährleisten. Hier stoßen derartige akzeptierte Beteiligungsverfahren schnell an Grenzen.

Auf die Vermittlungsproblematik wurde in Diskussionen zum Technikvertrauen in ähnlicher Weise hingewiesen. So komme es nicht nur darauf an, rein zufällig zu einer Deckung von individuellem Vertrauen und einem (unabhängig gefällten) ethischen Urteil zu gelangen, sondern gerade auch darauf, dass das Wissen um gute Gründe eben dieses Vertrauen (und damit eben auch Akzeptanz und Nutzung) motivieren soll (Nickel 2011, S. 357 f.). Hier sind zwar die rationalen Gründe der Vertrauens*würdigkeit* angesprochen, allerdings aus der Perspektive der Laien, nicht der Experten. Daher muss, wo immer öffentliches Vertrauen produktiv eine legitimierende Wirkung entfalten soll, die Vertrauenswürdigkeit von Technik anders vermittelt werden, als über eine technisch-wissenschaftliche Einschätzung der Verlässlichkeit bzw. der IT-Sicherheit. Entsprechend ist (für sich genommen) die öffentliche Verfügbarkeit des Quellcodes einer Software auch nicht automatisch ein guter Grund für öffentliches Technikvertrauen, denn für Laien besteht hier eine nur theoretische Möglichkeit der Prüfung, die wiederum ein entsprechendes Expertenwissen voraussetzt. Soll die Rückbindung an tech-

nische Expertise aber nicht gänzlich aufgegeben werden, bedarf es einer vermittelten Vertrauensbeziehung – was aber gerade nicht bedeutet, *Technikwissen* an Laien zu vermitteln, sondern *gute Gründe für Vertrauen* verständlich und nachvollziehbar zu machen (und dann ggf. auch für Laien eine selbstbestimmtere Teilhabe einzulösen). Welche Voraussetzungen gibt es für das Gelingen einer solchen Vermittlung von Vertrauen?

Bei vermitteltem, rational begründetem Vertrauen müssen mindestens zwei Vertrauensbeziehungen ineinandergreifen: 1) zwischen Expertin und einer durch sie als ausreichend verlässlich (*reliability*) und sicher (*security*) beurteilten Technik, sodann 2) aber noch zwischen Expertin und Anwendern bzw. Personen aus der Öffentlichkeit, die dem Urteil der Expertin vertrauen. Hierbei fällt auf, dass das, was letztlich von Laien anvertraut wird, über die Vermittlung hinweg durchgängig erhalten bleiben muss: Eine Expertise zu De-Mail müsste die Erwartungshaltung des Anwenders (z. B. die verlässliche, vertrauliche und manipulationsgeschützte Übermittlung von Unterlagen) abdecken. Dies muss für den Anwender zudem verständlich sein, denn nur dann lässt sich dessen Erwartungshaltung rational auf das Urteil der Expertin stützen. Allgemeiner ausgedrückt: Es geht darum, dass der Anwender die Expertise auf die konkreten Anforderungen seiner eigenen Situation beziehen kann. Was der Anwender von einer verlässlichen und sicheren Digitaltechnik erwartet, muss bei stimmiger Vermittlung des Vertrauens letztlich durch das ausgedrückt werden, woraufhin die Expertin die Technik bewertet.

Im Vertrauen, das ein Anwender in die Expertin setzt, bezieht sich der Erwartungshorizont des Vertrauens vermittelt auf die *Techniknutzung*, dennoch findet die *Begründung* der Vertrauenswürdigkeit der Expertin im zwischenmenschlichen Kontext statt; sie muss sich anhand zwischenmenschlicher Kriterien als vertrauenswürdig erweisen. Hier ist es für eine Technikbewertung möglich, auf die reichhaltige philosophische Kerndebatte zum Vertrauen zurückzugreifen, die den Rahmen dieses Artikels überschreitet. Als Kriterien für die Vertrauenswürdigkeit in funktionalen Beziehungen wird u. a. auf charakterliche Eigenschaften wie Kompetenz, Ehrlichkeit und Zuverlässigkeit verwiesen, die im unpersönlichen Kontext aber nur schwer zu bewerten sind und daher oft als entsprechende Reputation gedacht werden (O'Neill 2018, S. 299, 295). Hierauf kann auch institutionell für technische Überwachungsvereine, Warentest oder technische Normungsgremien rekurriert werden (Ropohl 2010, S. 129 f.).

Einen für den öffentlichen Kontext sehr interessanten Entwurf zur Begründung von Vertrauenswürdigkeit hat Philip Pettit (1995) vorgestellt: Weil öffentlich bekundetes Vertrauen immer auch Anerkennung und Wertschätzung ausdrückt, werden Vertrauensnehmer (etwa die Expertin) vermeiden wollen, dieses Vertrauen öffentlich zu enttäuschen. Für Pettit (1995, S. 203, 219) drückt diese Dynamik eine List des Vertrauens aus, die zu erklären vermag, warum Vertrauen unter bestimmten Umständen Vertrauenswürdigkeit zu stiften vermag, wir also nicht immer auf eine entsprechende Reputation verweisen müssen.

Derartige, auch auf den Laienkontext beziehbare Vertrauenskonzepte ermöglichen es, bei der normativen Technikgestaltung oder in partizipativen TA-Verfahren das Vermittlungsproblem zwischen Akzeptanz und Akzeptabilität stärker in den Vordergrund zu rücken. So ließen sich z. B. analog zur akzeptanzorientierten Technikgestaltung schon beim Entwurf digitaler Techniken gute Gründe für das Vertrauen durch Laien mitdenken, was gerade für kritische digitale Technikanwendungen wichtig wäre. Als Beispiel hierfür lässt sich ein Konzept für *privacy*-schonende Videoüberwachung nennen, bei dem eine Entschlüsselung von personenbezogenen Bildinformationen über Ombudspersonen geregelt, das Missbrauchspotentials dadurch reduziert und die Vertrauenswürdigkeit der Lösung nachvollziehbar gestärkt wird (Weydner-Volkmann und Feiten 2019).

Vertrauenskonzepte können das Vermittlungsproblem zwischen Akzeptanz und Akzeptabilität stärker in den Vordergrund rücken.

Fazit

‚Technikvertrauen' scheint in mindestens zweierlei Hinsicht einen wichtigen produktiven Beitrag für eine Technikbewertung leisten zu können: Zum einen erlaubt der Begriff einen expliziten Anschluss an die IT-Sicherheitsforschung und somit an einen Bereich, der gerade für Digitaltechniken von zentraler Bedeutung ist. Diese Aspekte gehen deutlich über Verlässlichkeitsfragen hinaus, die aktuell noch fast ausschließlich im Fokus der Technikbewertung stehen. Zum anderen bleibt insbesondere bei Digitaltechniken ein Vermittlungsproblem, das über die Doppelung von ‚Akzeptanz' und ‚Akzeptabilität' nur unzureichend adressiert wird: Auch für Laien muss sich Vertrauen z. B. in digitale Infrastrukturen *rational begründen* lassen, und zwar selbst dann, wenn staatliche Akteure und Hersteller zu potenziellen Angreifern gehören. Weil Technikvertrauen im Sinne eines Erwartungshorizonts auch *vermittelt* über Expertinnen und Institutionen konzipiert werden kann, erlaubt der Begriff der Vertrauenswürdigkeit, die Perspektive von Laien bereits in der Technikgestaltung begrifflich stimmig einzubinden.

Dank

Ich danke Alfred Nordmann und Klaus Kornwachs für ihre hilfreichen Kommentare.

Literatur

AI HLEG (2019): Ethics guidelines for trustworthy AI. Brüssel: High Level Expert Group on Artificial Intelligence. Online verfügbar unter https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, zuletzt geprüft am 20. 04. 2021.

Budnik, Christian (2016): Gründe für Vertrauen, Vertrauenswürdigkeit und Kompetenz. In: Deutsche Zeitschrift für Philosophie 64 (1), S. 103–118. https://doi.org/10.1515/dzph-2016-0007

Cap, Clemens (2015): Kann man einem Computer vertrauen? In: Josette Baer und Wolfgang Rother (Hg.): Vertrauen. Basel: Schwabe, S. 109–126.

Eusgeld, Irene; Fechner, Bernhard; Salfner, Felix; Walter, Max; Limbourg, Philipp; Zhang, Lijun (2008): Hardware reliability. In: Irene Eusgeld, Felix Freiling und Ralf Reussner (Hg.): Dependability metrics. Berlin: Springer, S. 59–103. https://doi.org/10.1007/978-3-540-68947-8_9

Gethmann, Carl Friedrich; Sander, Thorsten (1999): Rechtfertigungsdiskurse. In: Armin Grunwald und Stephan Saupe (Hg.): Ethik in der Technikgestaltung. Praktische Relevanz und Legitimation. Berlin: Springer, S. 117–151. https://doi.org/10.1007/978-3-642-60033-3_7

GlobalPlatform (2018): Root of trust definitions and requirements. Online verfügbar unter https://globalplatform.org/wp-content/uploads/2018/07/GP_RoT_Definitions_and_Requirements_v1.1_PublicRelease-2018-06-28.pdf, zuletzt geprüft am 20. 04. 2021.

Grunwald, Armin (2005): Zur Rolle von Akzeptanz und Akzeptabilität von Technik bei der Bewältigung von Technikkonflikten. In: TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 14 (3), S. 54–60. https://doi.org/10.14512/tatup.14.3.54

Hartmann, Martin (2010): Die Komplexität des Vertrauens. In: Matthias Maring (Hg.): Vertrauen. Zwischen sozialem Kitt und der Senkung von Transaktionskosten. Karlsruhe: KIT Scientific Publishing, S. 15–25. https://doi.org/10.26530/OAPEN_422381

Jaufmann, Dieter (1999): Technikakzeptanzforschung. In: Stephan Bröchler, Georg Simonis und Karsten Sundermann (Hg.): Handbuch Technikfolgenabschätzung. Berlin: Edition Sigma, S. 205–226.

Jones, Karen (2020): Trust. In: International Encyclopedia of Ethics. Malden, MA: Wiley-Blackwell, S. 1–9. https://doi.org/10.1002/9781444367072.wbiee665.pub2

Jones, Karen (2012): Trustworthiness. In: Ethics 123 (1), S. 61–85. https://doi.org/10.1086/667838

Kaminski, Andreas (2010): Technik als Erwartung. Grundzüge einer allgemeinen Technikphilosophie. Bielefeld: transcript. https://doi.org/10.14361/transcript.9783839414705

Köhl, Harald (2001): Vertrauen als zentraler Moralbegriff? In: Martin Hartmann und Claus Offe (Hg.): Vertrauen. Die Grundlage des sozialen Zusammenhalts. Frankfurt a. M.: Campus, S. 114–140.

Lampe, Hildrun; Kaminski, Andreas (2019): Verlässlichkeit und Vertrauenswürdigkeit von Computersimulationen. In: Kevin Liggieri und Oliver Müller-Känel (Hg.): Mensch-Maschine-Interaktion. Handbuch zu Geschichte – Kultur – Ethik. Berlin: J. B. Metzler, S. 325–331. https://doi.org/10.1007/978-3-476-05604-7_60

Metzinger, Thomas (2019): Ethik-Waschmaschinen made in Europe. In: Tagesspiegel.de. Online verfügbar unter https://background.tagesspiegel.de/ethik-waschmaschinen-made-in-europe, zuletzt geprüft am 20. 04. 2021.

Nickel, Philip (2011): Ethics in e-trust and e-trustworthiness. The case of direct computer-patient interfaces. In: Ethics and Information Technology 13 (4), S. 355–363. https://doi.org/10.1007/s10676-011-9271-9

O'Neill, Onora (2018): Linking trust to trustworthiness. In: International Journal of Philosophical Studies 26 (2), S. 293–300. https://doi.org/10.1080/09672559.2018.1454637

Pettit, Philip (1995): The cunning of trust. In: Philosophy and Public Affairs 24 (3), S. 202–225. https://doi.org/10.1111/j.1088-4963.1995.tb00029.x

Ropohl, Günter (2010): Das Misstrauen in der Technikdebatte. In: Matthias Maring (Hg.): Vertrauen. Zwischen sozialem Kitt und der Senkung von Transaktionskosten. Karlsruhe: KIT Scientific Publishing, S. 115–132.

Weber, Arnd et al. (2020 a): Sichere IT ohne Schwachstellen und Hintertüren. In: TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 29 (1), S. 30–36. https://doi.org/10.14512/tatup.29.1.30

Weber, Karsten; Christen, Markus; Herrmann, Dominik (2020 b): Bedrohung, Verwundbarkeit, Werte und Schaden. Cyberattacken und Cybersicherheit als Thema der Technikfolgenabschätzung. In: TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 29 (1), S. 11–15. https://doi.org/10.14512/tatup.29.1.11

Weydner-Volkmann, Sebastian (2018): Moralische Landkarten der Sicherheit. Ein Framework zur hermeneutisch-ethischen Bewertung von Fluggastkontrollen im Anschluss an John Dewey. Baden-Baden: Ergon Verlag. https://doi.org/10.5771/9783956503788

Weydner-Volkmann, Sebastian; Feiten, Linus (2019): Vertrauensstiftende Videoüberwachung? In: digma. Zeitschrift für Datenrecht und Informationssicherheit 19 (4), S. 218–221.

PROF. DR. SEBASTIAN WEYDNER-VOLKMANN

ist Juniorprofessor für Ethik der digitalen Methoden und Techniken an der Ruhr-Universität Bochum. Nach dem Studium in Freiburg promovierte er 2017 am dortigen Centre for Security and Society sowie am Husserl-Archiv im Schnittbereich von Technikethik, Moralpragmatismus und Sicherheitsforschung.